Netflix _DataAnalysis Report

December 6, 2023

1 Netflix Data Analysis

This dataset consists of TV shows and movies available on Netflix as of 2019. The dataset is collected from Kaggle.com. In this notebook we are going to process and analyze the data using Numpy and Pandas libraries and then we will visualize it using matplotlib library.

```python
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
```

```python
data = pd.read_csv('netflix_dataset.csv')
```

```python
data.head()
```

```
     show_id      type                                       title  \
0   81145628     Movie  Norm of the North: King Sized Adventure
1   80117401     Movie              Jandino: Whatever it Takes
2   70234439   TV Show                        Transformers Prime
3   80058654   TV Show        Transformers: Robots in Disguise
4   80125979     Movie                              #realityhigh

                 director  \
0  Richard Finn, Tim Maltby
1                      NaN
2                      NaN
3                      NaN
4         Fernando Lebrija

                                       cast  \
0  Alan Marriott, Andrew Toth, Brian Dobson, Cole…
1                            Jandino Asporaat
2  Peter Cullen, Sumalee Montano, Frank Welker, J…
3  Will Friedle, Darren Criss, Constance Zimmer, …
4  Nesta Cooper, Kate Walsh, John Michael Higgins…

                              country        date_added  release_year  \
0  United States, India, South Korea, China  September 9, 2019          2019
```

```
1                     United Kingdom  September 9, 2016          2016
2                     United States   September 8, 2018          2013
3                     United States   September 8, 2018          2016
4                     United States   September 8, 2017          2017

    rating  duration                           listed_in  \
0    TV-PG    90 min  Children & Family Movies, Comedies
1    TV-MA    94 min                      Stand-Up Comedy
2  TV-Y7-FV  1 Season                             Kids' TV
3    TV-Y7  1 Season                             Kids' TV
4    TV-14    99 min                             Comedies

                                         description
0  Before planning an awesome wedding for his gra…
1  Jandino Asporaat riffs on the challenges of ra…
2  With the help of three human allies, the Autob…
3  When a prison ship crash unleashes hundreds of…
4  When nerdy high schooler Dani finally attracts…
```

[ ]: `data.shape`

[ ]: (6234, 12)

[ ]: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6234 entries, 0 to 6233
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   show_id       6234 non-null   int64
 1   type          6234 non-null   object
 2   title         6234 non-null   object
 3   director      4265 non-null   object
 4   cast          5664 non-null   object
 5   country       5758 non-null   object
 6   date_added    6223 non-null   object
 7   release_year  6234 non-null   int64
 8   rating        6224 non-null   object
 9   duration      6234 non-null   object
 10  listed_in     6234 non-null   object
 11  description   6234 non-null   object
dtypes: int64(2), object(10)
memory usage: 584.6+ KB
```

[ ]: `data.columns`

```
[ ]: Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in', 'description'],
      dtype='object')
```

```
[ ]: data_copy = data.copy()
```

```
[ ]: data_copy.head()
```

```
[ ]:    show_id     type                                 title  \
     0  81145628    Movie  Norm of the North: King Sized Adventure
     1  80117401    Movie                Jandino: Whatever it Takes
     2  70234439  TV Show                         Transformers Prime
     3  80058654  TV Show        Transformers: Robots in Disguise
     4  80125979    Movie                           #realityhigh

                    director  \
     0  Richard Finn, Tim Maltby
     1                     NaN
     2                     NaN
     3                     NaN
     4         Fernando Lebrija

                                        cast  \
     0  Alan Marriott, Andrew Toth, Brian Dobson, Cole…
     1                            Jandino Asporaat
     2  Peter Cullen, Sumalee Montano, Frank Welker, J…
     3  Will Friedle, Darren Criss, Constance Zimmer, …
     4  Nesta Cooper, Kate Walsh, John Michael Higgins…

                                    country        date_added  release_year  \
     0  United States, India, South Korea, China  September 9, 2019          2019
     1                            United Kingdom  September 9, 2016          2016
     2                             United States  September 8, 2018          2013
     3                             United States  September 8, 2018          2016
     4                             United States  September 8, 2017          2017

           rating  duration                          listed_in  \
     0      TV-PG    90 min  Children & Family Movies, Comedies
     1      TV-MA    94 min                       Stand-Up Comedy
     2   TV-Y7-FV  1 Season                           Kids' TV
     3      TV-Y7  1 Season                           Kids' TV
     4      TV-14    99 min                           Comedies

                                    description
     0  Before planning an awesome wedding for his gra…
     1  Jandino Asporaat riffs on the challenges of ra…
     2  With the help of three human allies, the Autob…
```

```
3   When a prison ship crash unleashes hundreds of…
4   When nerdy high schooler Dani finally attracts…
```

```
[ ]: data.drop(['director','cast','date_added'],axis=1, inplace=True)
```

```
[ ]: data.head()
```

```
[ ]:      show_id     type                                      title  \
     0  81145628    Movie  Norm of the North: King Sized Adventure
     1  80117401    Movie               Jandino: Whatever it Takes
     2  70234439  TV Show                         Transformers Prime
     3  80058654  TV Show          Transformers: Robots in Disguise
     4  80125979    Movie                              #realityhigh

                                 country  release_year   rating  duration  \
     0  United States, India, South Korea, China          2019    TV-PG    90 min
     1                          United Kingdom            2016    TV-MA    94 min
     2                           United States            2013  TV-Y7-FV  1 Season
     3                           United States            2016    TV-Y7   1 Season
     4                           United States            2017    TV-14    99 min

                              listed_in  \
     0  Children & Family Movies, Comedies
     1                    Stand-Up Comedy
     2                          Kids' TV
     3                          Kids' TV
     4                           Comedies

                                    description
     0  Before planning an awesome wedding for his gra…
     1  Jandino Asporaat riffs on the challenges of ra…
     2  With the help of three human allies, the Autob…
     3  When a prison ship crash unleashes hundreds of…
     4  When nerdy high schooler Dani finally attracts…
```

```
[ ]: data.isnull().sum()
```

```
[ ]: show_id           0
     type              0
     title             0
     country         476
     release_year      0
     rating           10
     duration          0
     listed_in         0
     description       0
     dtype: int64
```

```
[ ]: data['country'].replace(np.nan, 'United States', inplace=True)
```

```
[ ]: data['rating'].replace(np.nan,'TV-MA', inplace=True)
```

```
[ ]: data.isnull().sum()
```

```
[ ]: show_id        0
     type           0
     title          0
     country        0
     release_year   0
     rating         0
     duration       0
     listed_in      0
     description    0
     dtype: int64
```

```
[ ]: data.head()
```

```
[ ]:      show_id      type                                   title  \
     0  81145628     Movie   Norm of the North: King Sized Adventure
     1  80117401     Movie                 Jandino: Whatever it Takes
     2  70234439   TV Show                          Transformers Prime
     3  80058654   TV Show         Transformers: Robots in Disguise
     4  80125979     Movie                              #realityhigh

                                   country  release_year    rating  duration  \
     0  United States, India, South Korea, China         2019     TV-PG     90 min
     1                          United Kingdom           2016     TV-MA     94 min
     2                           United States          2013  TV-Y7-FV  1 Season
     3                           United States          2016     TV-Y7  1 Season
     4                           United States          2017     TV-14     99 min

                             listed_in  \
     0  Children & Family Movies, Comedies
     1                     Stand-Up Comedy
     2                           Kids' TV
     3                           Kids' TV
     4                           Comedies

                                 description
     0  Before planning an awesome wedding for his gra…
     1  Jandino Asporaat riffs on the challenges of ra…
     2  With the help of three human allies, the Autob…
     3  When a prison ship crash unleashes hundreds of…
     4  When nerdy high schooler Dani finally attracts…
```
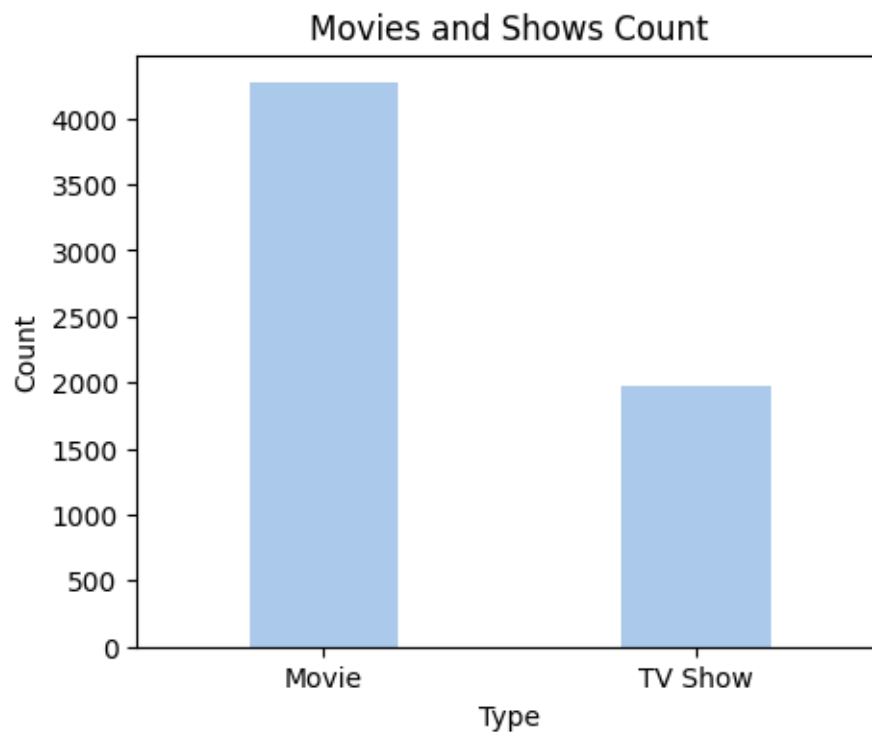
Q. What is the movie and TV show count?

```
[ ]: data["type"].value_counts()
```

```
[ ]: type
     Movie      4265
     TV Show    1969
     Name: count, dtype: int64
```

```
[ ]: sns.set_palette("pastel")
     plt.figure(figsize=(5,4))
     sns.countplot(x="type", data=data, width=0.4)
     plt.title('Movies and Shows Count')
     plt.xlabel('Type')
     plt.ylabel('Count')
     plt.show()
```


Hence, from the plot, we can come to the conclusion, that total number of movies in the dataset is 4265 and total number of TV shows in the dataset is 1969.

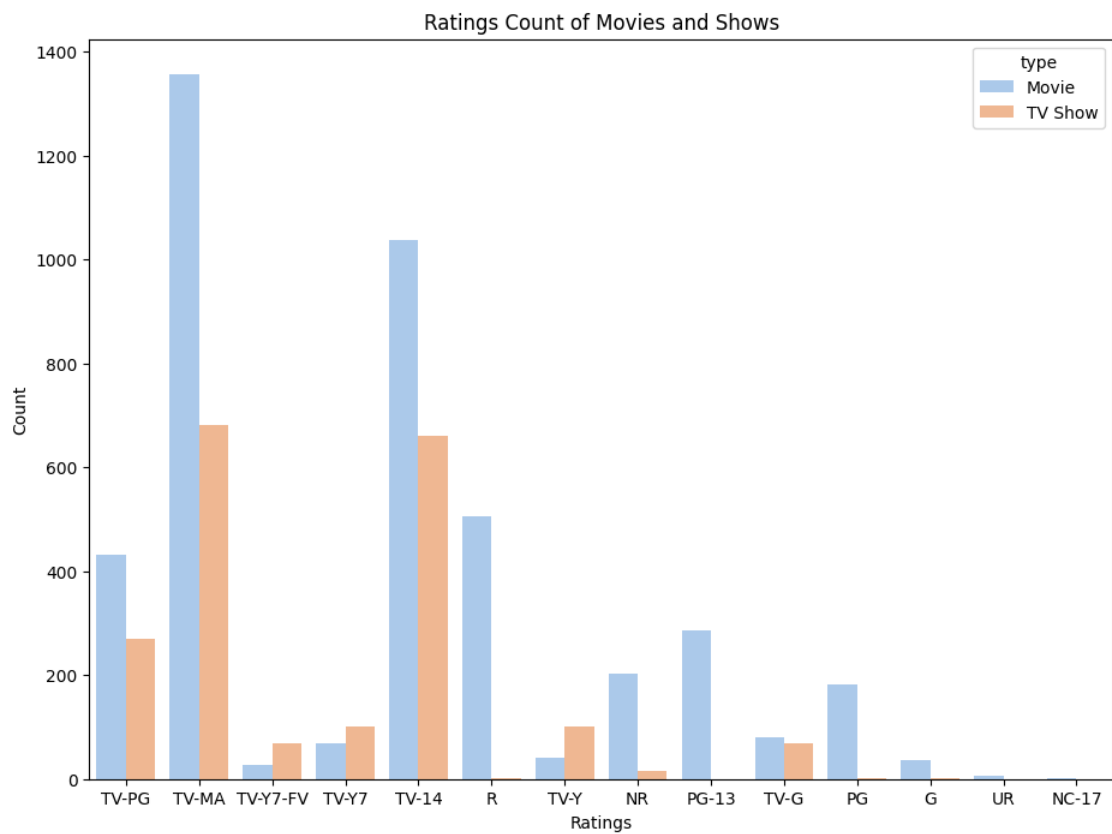Q. Rating count of Movies and TV Shows.

```
[ ]: data["rating"].value_counts()
```

```
[ ]: rating
     TV-MA        2037
```

```
TV-14          1698
TV-PG           701
R               508
PG-13           286
NR              218
PG              184
TV-Y7           169
TV-G            149
TV-Y            143
TV-Y7-FV         95
G                37
UR                7
NC-17             2
Name: count, dtype: int64
```

```python
plt.figure(figsize=(11,8))
plt.title('Ratings Count of Movies and Shows')
plt.xlabel('Ratings')
plt.ylabel('Count')
sns.countplot(x="rating",hue='type', data=data)
plt.show()
```
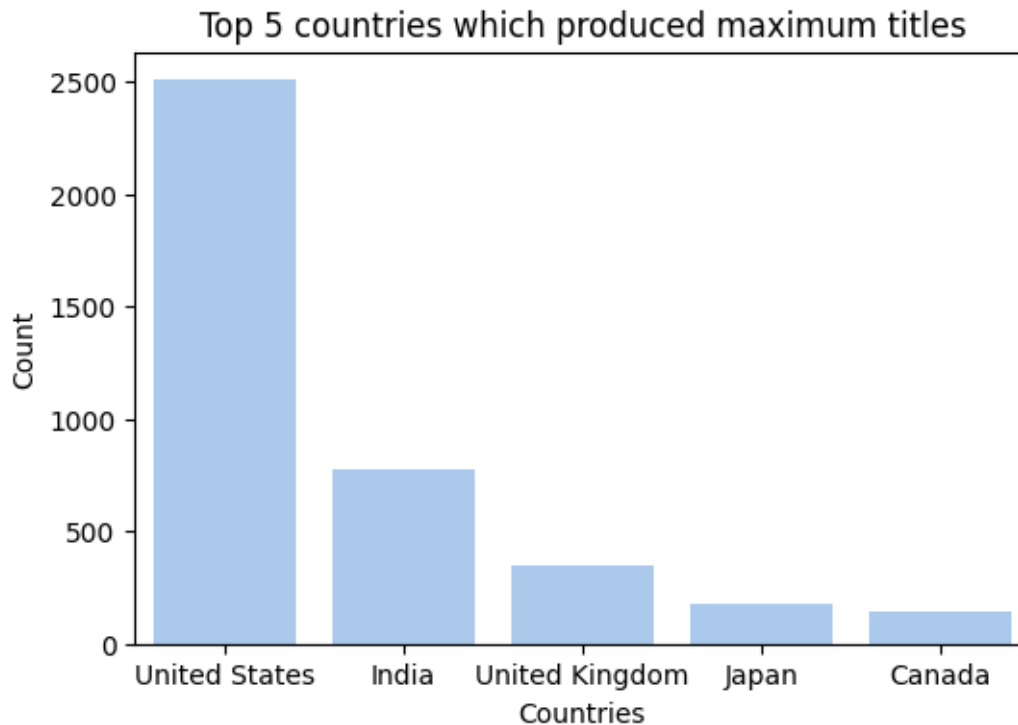
Hence, from the plot, we can tell the rating count of movies and TV shows in the dataset.

Q. What are the top 5 countries with maximum releases?

```
[ ]: datatemp = pd.DataFrame(data["country"].value_counts().head())
     datatemp
```

```
[ ]:                    count
     country
     United States      2508
     India               777
     United Kingdom      348
     Japan               176
     Canada              141
```

```
[ ]: # MAx Releases
     plt.figure(figsize=(6,4))
     sns.barplot(x="country", y="count", data=datatemp)
     plt.title('Top 5 countries which produced maximum titles')
     plt.xlabel('Countries')
     plt.ylabel('Count')
     plt.show()
```


Hence, from the plot, we can tell that the maximum releases were by United States, followed by
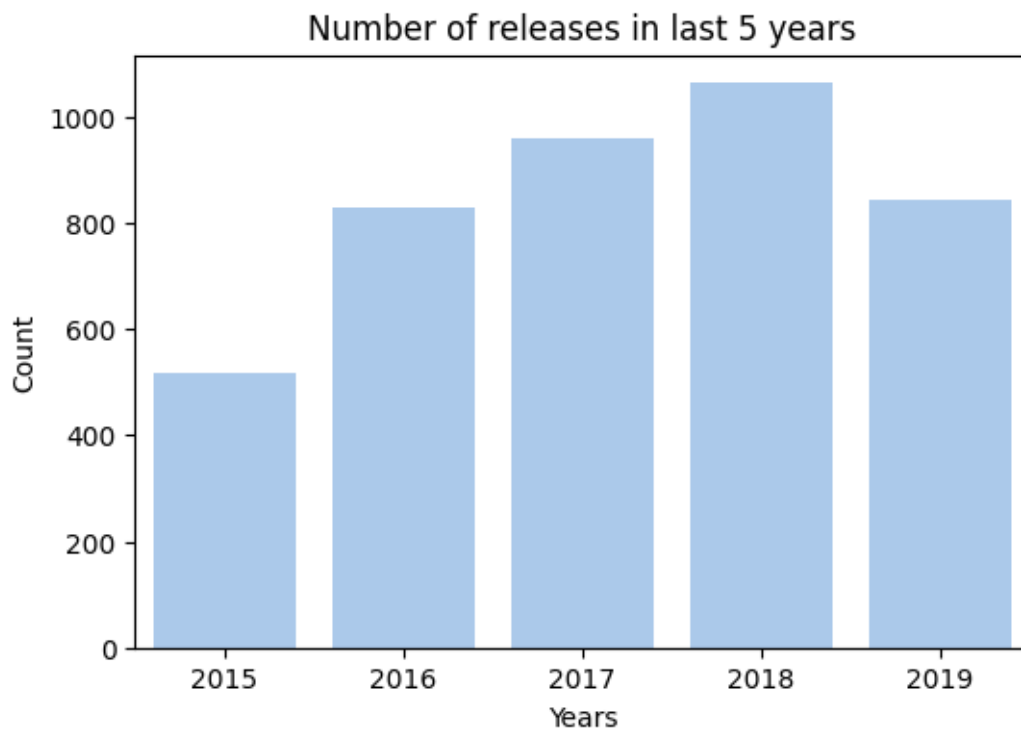
India, United Kingdom, Japan and Canada.

Q. What are the number of releases in last 5 years?

```
[ ]: datatemp1 = pd.DataFrame(data["release_year"].value_counts().head())
     datatemp1
```

```
[ ]:                 count
     release_year
     2018            1063
     2017             959
     2019             843
     2016             830
     2015             517
```

```
[ ]: plt.figure(figsize=(6,4))
     sns.barplot(x="release_year", y="count", data=datatemp1)
     plt.title('Number of releases in last 5 years')
     plt.xlabel('Years')
     plt.ylabel('Count')
     plt.show()
```
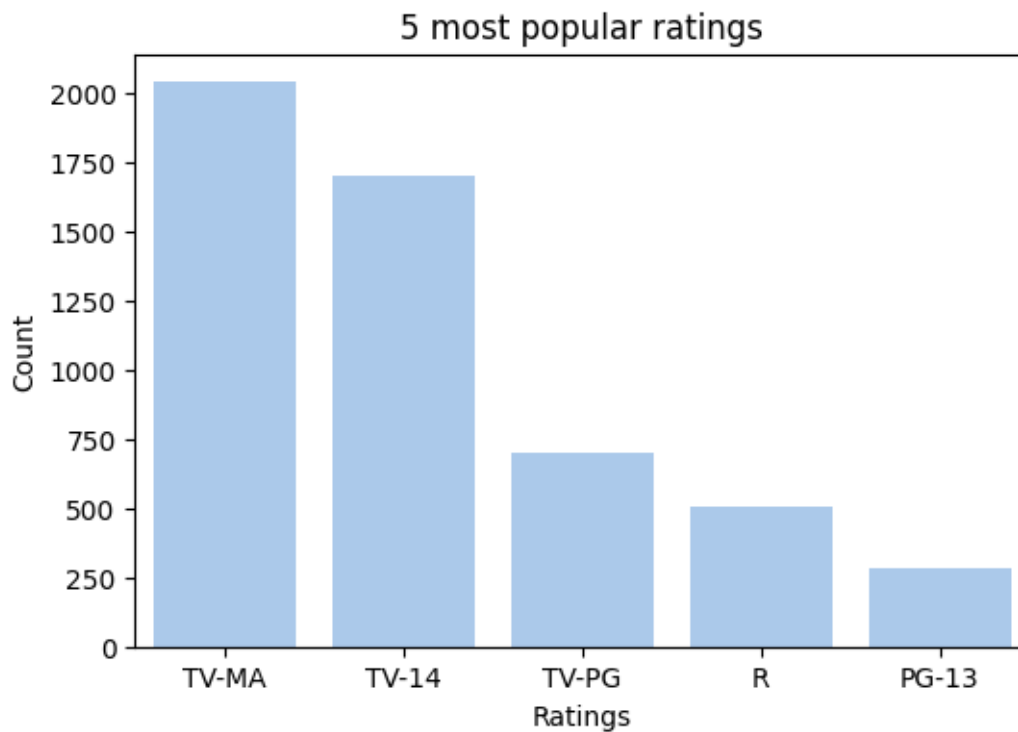

Hence, from the plot, we can tell that the maximum releases were in the year 2018, followed by 2017,2019, 2016 and 2015.

Q. What are the top 5 most popular ratings?

```
[ ]: datatemp2 = pd.DataFrame(data["rating"].value_counts()).head()
     datatemp2
```

```
[ ]:            count
     rating
     TV-MA       2037
     TV-14       1698
     TV-PG        701
     R            508
     PG-13        286
```

```
[ ]: plt.figure(figsize=(6,4))
     sns.barplot(x="rating",y="count", data=
     datatemp2)
     plt.title('5 most popular ratings')
     plt.xlabel('Ratings')
     plt.ylabel('Count')
     plt.show()
```
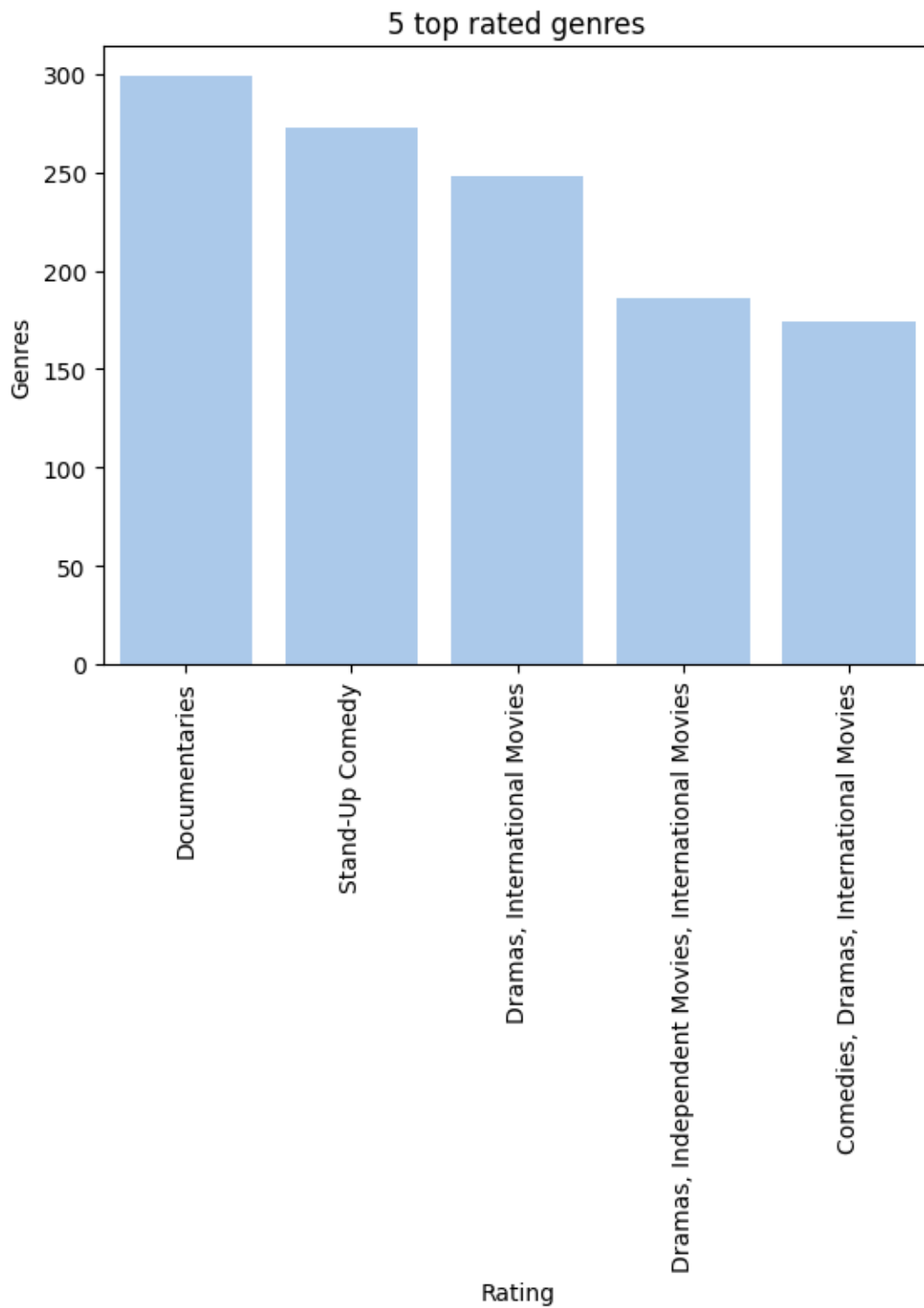

Hence, from the plot, we can tell that the 5 most popular ratings were, TV-MA, TV-14, TV-PG, R, PG-13

Q. What are the top 5 most popular genres?

```
datatemp3 = pd.DataFrame(data["listed_in"].value_counts().head())
datatemp3
```

	count
listed_in	
Documentaries	299
Stand-Up Comedy	273
Dramas, International Movies	248
Dramas, Independent Movies, International Movies	186
Comedies, Dramas, International Movies	174

```
plot = sns.barplot(x="listed_in",y="count",data=datatemp3)
plt.setp(plot.get_xticklabels(), rotation=90)
plt.title('5 top rated genres')
plt.xlabel('Rating')
plt.ylabel('Genres')
plt.show()
```

5 top rated genres

Hence, from the plot, we can tell that the maximum releases were in Documentaries genre, followed by Stand-Up comedy, then Dramas, International movie,etc.

```
datatemp4 = pd.DataFrame(data["listed_in"].value_counts().tail())
datatemp4
```

```
                                                  count
listed_in
TV Comedies, TV Dramas, TV Mysteries                  1
Classic Movies, Dramas, LGBTQ Movies                  1
Crime TV Shows, TV Dramas, TV Sci-Fi & Fantasy        1
Docuseries, Reality TV, Teen TV Shows                 1
Reality TV, Spanish-Language TV Shows                 1
```

For the least rated genres, it is evident that all of them have releases as 1, so no point plotting them.